Cybersecurity and Law

2025 Nr 1(13)

DOI: 10.35467/cal/214535

Deepfake as a threat and challenge to national information security

Deepfake jako zagrożenie i wyzwanie dla bezpieczeństwa informacyjnego państwa

Piotr GALEMBA
Akademia Sztuki Wojennej / War Studies University
ORCID: 0000-0003-0551-4185
E-mail: piotr.galemba@gmail.com

Abstract

The phenomenon of deepfakes, understood as the generation of extremely realistic but completely fake audiovisual materials based on imported input data, due to its potential, apart from positive purposes (such as entertainment or education), can also be used as a tool for information, disinformation, and psychological operations, thus creating new and exacerbating existing threats to national information security. The aim of this article is to systematize knowledge about the essence of the deepfake phenomenon and to identify the main threats and challenges to national information security resulting from it.

The above objective was achieved and the question formulated in the research problem was answered by conducting research using the method of analysis and criticism of literature.

The research indicated that the main threats to state information security are the discrediting of commanders and political leadership, lowering the morale of the military and society, increasing social tensions, and increasing information chaos. At the same time, with the increasing use of this technology, states face new challenges, including in the areas of public education, legal regulation, and the ability to verify and detect false images and recordings.

Due to the fact that deepfake technology is closely linked to the development of AI and the ever-increasing possibilities offered by the exchange of information on the

Internet, it is necessary to take appropriate steps to explore its impact on the information sphere, while also taking steps to introduce appropriate legal regulations and increase society's resilience to the potential threats arising from its use.

Keywords

deepfake, information security, artificial intelligence, disinformation, propaganda

Streszczenie

Zjawisko *deepfake*, rozumiane jako generowanie niezwykle realistycznych, ale całkowicie fałszywych materiałów audiowizualnych opartych na zaimportowanych danych wejściowych, z uwagi na swój potencjał oprócz celów pozytywnych (jak rozrywka, czy edukacja), mogą być wykorzystywane również jako narzędzie prowadzonych działań informacyjnych, dezinformacyjnych oraz psychologicznych, stwarzając tym samym nowe oraz potęgując dotychczas znane zagrożenia dla bezpieczeństwa informacyjnego państwa. Celem niniejszego artykułu jest usystematyzowanie wiedzy na temat istoty zjawiska *deepfake* oraz wskazanie wynikających z niego głównych zagrożeń i wyzwań dla bezpieczeństwa informacyjnego państwa w XXI wieku.

Realizacja powyższego celu oraz udzielenie odpowiedzi na sformułowane w problemie badawczym pytanie, możliwe było poprzez przeprowadzenie badań przy wykorzystaniu metody analizy i krytyki piśmiennictwa.

W toku przeprowadzonych badań wskazano, że głównymi zagrożeniami dla bezpieczeństwa informacyjnego państwa są dyskredytowanie dowódców i kierownictwa politycznego, obniżanie morale wojska i społeczeństwa, potęgowania napięć społecznych, oraz zwiększenie chaosu informacyjnego. Jednocześnie wraz ze wzrostem wykorzystywania tej technologii, państwa stają przed nowymi wyzwaniami, m.in. w zakresie edukacji społeczeństwa, regulacji prawnych, czy też możliwości weryfikacji i wykrywania fałszywych obrazów i nagrań.

Z uwagi na fakt, że technologia *deepfake*, ściśle powiązana jest z rozwojem sztucznej inteligencji oraz stale wzrastającymi możliwościami jakie daje wymiana informacji w Internecie, konieczne jest podjęcie odpowiednich kroków w zakresie zgłębiania jego wpływu na sferę informacyjną, z jednoczesnym podjęciem kroków w zakresie wprowadzenia stosownych regulacji prawnych oraz zwiększanie odporności społeczeństwa na potencjalne zagrożenia wynikające z ich stosowania.

Słowa kluczowe

deepfake, bezpieczeństwo informacyjne, sztuczna inteligencja, dezinformacja, propaganda

This article was written as part of an international project entitled „Protection of Image in Copyright Law in the Digital Media Era - A Polish-Spanish Comparative Approach" within the framework of cooperation between the Faculty of Law and Administration at the War Studies University in Warsaw and the Faculty of Law at Pablo de Olavide University in Seville, Spain.

Wprowadzenie

Dostęp do rzetelnych oraz zweryfikowanych informacji stanowi jeden z kluczowych determinantów przetrwania oraz niezakłóconego rozwoju każdego podmiotu, zaczynając od jednostki, poprzez grupy społeczne, kończąc na państwach oraz organizacjach międzynarodowych. Szczególnie w dzisiejszym zglobalizowanym świecie, ściśle powiązanym z permanentnym postępem technologicznym oraz zwiększającą się ilością rozpowszechnianych i przetwarzanych danych i informacji nabiera to coraz większego znaczenia, stwarzając jednocześnie nowe wyzwania i zagrożenia dla bezpieczeństwa informacyjnego. Jednymi z nich jest rozpowszechnianie częściowo zafałszowanych lub całkowicie spreparowanych nieprawdziwych informacji, w celu oddziaływania psychologicznego oraz informacyjnego na określone grupy odbiorców. O ile zjawisko dezinformacji i propagandy nie jest niczym nowym, a ich wykorzystywanie można odnaleźć w historii człowieka niemal od początku jego działalności, o tyle współczesne rozwiązania techniczne oraz technologiczne stwarzają nowe możliwości oraz zakres kreowania treści wykorzystywanych na potrzeby prowadzenia tych działań.

W sposób szczególny wpływ na to ma rozwój sztucznej inteligencji, która pozwala nie tylko na skrócenie czasu przygotowania, czy też rozpowszechniania wygenerowanych treści, ale odgrywa coraz większą rolę w zakresie generowania treści audiowizualnych, które mimo wytworzenie w całości przez programy komputerowe, dzięki swojej dbałości o szczegóły, są niemal nie do odróżnienia względem świata rzeczywistego, określane mianem *deepfake*.

Celem niniejszego artykułu jest usystematyzowanie wiedzy na temat istoty zjawiska *deepfake* oraz wskazanie wynikających z niego głównych zagrożeń oraz wyzwań dla bezpieczeństwa informacyjnego państwa w XXI wieku. Realizacja tego celu jest możliwa poprzez udzielenie odpowiedzi na następujący problem badawczy: Jakie zagrożenia i wyzwania dla bezpieczeństwa informacyjnego państwa wynikają z technologii *deepfake*?

Pojęcie bezpieczeństwa informacyjnego

Prowadząc rozważania poświęcone wpływu zagrożeń bezpieczeństwa informacyjnego wynikających ze stosowania *deepfake*, konieczne jest właściwe zrozumienie oraz charakterystyka samego pojęcia bezpieczeństwo informacyjne, które z uwagi na swój charakter, rozpatrywany jest wieloaspektowo, o czym świadczy fakt, że w literaturze przedmiotu można odnaleźć wiele odmiennych sposobów definiowania. Jednym z przykładów jest definicja zaproponowana przez autorów *Słownika terminów z zakresu bezpieczeństwa*, w której określone zostało, że „dotyczy informacji, we wszystkich etapach ich wytwarzania, przetwarzania, przechowywania i przysyłania, realizowane poprzez

przeciwdziałanie przed bezprawnym dostępem i jakąkolwiek ingerencję w dane, informacje i systemy informacyjne”[1].

Inna definicja została przedstawiona w Leksykonie cyberbezpieczeństwa, zgodnie z którą „bezpieczeństwo informacyjne państwa stanowi transsektorową dziedzinę bezpieczeństwa narodowego, będącą procesem polegającym na dążeniu do zapewnienia wolnego od zakłóceń funkcjonowania i rozwoju państwa w przestrzeni informacyjnej[2].

W rozumieniu E. Nowaka i M. Nowaka, rozpatrywać należy je w kontekście możliwości przetrwania i rozwoju społeczeństwa informacyjnego[3], natomiast zdaniem R. Kwećki, jest to „proces (lub jego chwilowy stan) realizowany (utrzymywany) w kooperacji pozytywnej, prowadzonej przez narzędzia będące w dyspozycji podmiotu i przeznaczone do działań w przestrzeni bezpieczeństwa informacyjnego”[4].

Na podstawie powyższych definicji można zatem wyciągnąć wniosek, że bezpieczeństwo informacyjne dotyczy zapewnienia bezpieczeństwa w zakresie przetwarzania oraz przekazywania informacji, a także zapobiegania ich modyfikacji i nieuprawnionemu wykorzystywaniu, które mogą wywołać negatywny wpływ na funkcjonowanie społeczeństwa informacyjnego lub podmiotów państwowych, na skutek oddziaływania na przestrzeń informacyjną państwa.

W kontekście omawianego w niniejszej pracy zagadnienia *deepfake*, wraz z rozwijającymi się możliwościami w zakresie kreowania oraz rozpowszechniania generowanych treści, z jednoczesnym ograniczeniem możliwości ich szybkiej i skutecznej weryfikacji, potęgowane są obecne oraz powstają nowe, dotychczas nieznane zagrożenia i wyzwania w zakresie możliwości jego zapewnienia, które dotyczą przede wszystkim sfery wiarygodności oraz modyfikowania danych i informacji w celu stworzenia chaosu informacyjnego oraz nowych podatności na zagrożenia, nie tylko w zakresie samego bezpieczeństwa informacyjnego, ale przenikając również do pozostałych sfer życia społecznego i wywoływania wpływu na bezpieczeństwo wewnętrzne, ekonomiczne, czy też militarne.

Istota *deepfake*

Manipulowanie percepcją odbiorcy wykorzystujące fotomontaże oraz falsyfikaty, nie jest nowym zjawiskiem, jednakże wraz z rozwojem technologicznym oraz zwiększeniem powszechnego dostępu do narzędzi opartych na sztucznej inteligencji intensywność oraz zakres ich wykorzystywania

[1] Słownik terminów z zakresu bezpieczeństwa, J. Pawłowski, B. Zdrodowski, M. Kuliczkowski (red.), Toruń 2020, s. 24.
[2] K. Chałubińska-Jentkiewicz, Bezpieczeństwo informacyjne, [w:] K. Chałubińska-Jentkiewicz (red.), Leksykon cyberbepieczeństwa, Warszawa 2024, s. 32.
[3] E. Nowak, M. Nowak, Zarys teorii bezpieczeństwa narodowego, Warszawa 2011, s. 103.
[4] R. Kwećka, Bezpieczeństwo informacyjne, [w:] J. Pawłowski (red.), Podstawy bezpieczeństwa narodowego (państwa), Warszawa 2017, s. 411.

znacząco się zwiększyły. Sam termin *deepfake*, jest stosunkowo nowy, bowiem po raz pierwszy to określenie pojawiło się w 2017 roku i było związane z nikiem użytkownika, który umieścił na platformie Reddit filmy pornograficzne wygenerowane poprzez wkomponowanie w oryginalne filmy nałożonych twarzy znanych aktorek (m.in. Emma Watson, czy Scarlet Johansson)[5]. Obecnie pod terminem *deepfake*, rozumie się „rodzaj sztucznej inteligencji do tworzenia będących oszustwem realistycznych obrazów rzeczywistości"[6] i odnosi się zarówno do algorytmów wykorzystywanych do kreowania tych treści, jak również do wygenerowanych produktów[7]. Należy w tym miejscu zadać zatem pytanie: Co odróżnia *deepfake* od znanych dotychczas falsyfikatów wykorzystywanych do celów propagandowych i dezinformacyjnych?

Udzielenie odpowiedzi na tak postawione pytanie wymaga bliższego omówienia sposobu tworzenia oraz charakterystycznych cech produktów *deepfake.* Istota tego zjawiska widoczna jest już w samej jego nazwie, która powstała poprzez połączenie dwóch angielskich słów *deep* (głęboki – w rozumieniu głębokiego uczenia) oraz *fake* (fałszywy, podrobiony). Przygotowywanie produktów *deepfake* polega na zaimportowaniu do specjalistycznego programu komputerowego odpowiednio wyselekcjonowanych danych wejściowych w formie próbek obrazów, nagrań głosowych, czy też nagrań wideo, które mają zostać poddane obróbce. Na ich podstawie algorytmy oparte o samouczącą się sztuczną inteligencję typuje charakterystyczne cechy wspólne (np. mimikę, gestykulację, ton głosu, etc.), na bazie których następnie prowadzone jest ich scalanie oraz kreowanie określonego materiału[8]. Poprzez zastosowanie dwóch algorytmów – generatora (odpowiedzialnego za filtrowanie i scalanie treści na podstawie cech wspólnych) oraz dyskryminatora (służącego do weryfikacji wygenerowanych treści), proces tworzenia trwa do chwili, gdy treści generowana oraz weryfikowana są ze sobą na tyle zgodne, że mogą być uznane przez system za spójne ze sobą[9].

Jednakże, w tym miejscu należy natomiast wspomnieć, że *deepfake* nie zawsze musi nosić znamiona działalności szkodliwej i nieetycznej. Zdaniem A. Piecucha klasyfikację *deepfake* można przedstawić, dzieląc je na sześć kategorii: rozrywkę, wirtualne postacie, *fake news*, pornografię, logowanie (autoryzację) oraz ataki finansowe[10].

---

[5] O. Wasiuta, S. Wasiuta, Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość, „Annales Univeristatis Paedagogicae Cracoviensis. Studia de Securitate", 2019, nr 9(3), s. 21.
[6] K. Kaczmarek, Deepfake, [w:] K. Chałubińska-Jentkiewicz (red.), Leksykon cyberbepieczeństwa, Warszawa 2024, s. 97.
[7] Zob.: O. Wasiuta, S. Wasiuta, op.cit., s. 30-21; K. M. Kiełpiński, Deepfake jako narzędzie do przekazywania informacji fałszywej i domniemanej. Analiza prawnokarna i cybernetyczna, „Kwartalnik Krajowej Szkoły Sądownictwa i Prokuratury", 2023, z. 3(51), s. 87, K. Kaczmarek, op.cit., s. 97.
[8] Porównaj: K. M. Kiełpiński, op.cit., s. 88 oraz K. Mularz, Deepfake jako generator fikcyjnych treści i rzeczywistej nieufności, „Dydaktyka Informatyki", 2024, nr 19, s. 57.
[9] Mularz K., op.cit., s. 57.
[10] A. Piecuch, Życie SMART - ma swoją cenę, „Dydaktyka Informatyki" 2020, nr 15, s. 61.

Według klasyfikacji zaprezentowanej przez I. Dąbrowską ze względu na cel jaki przyświeca twórcy, można podzielić wygenerowane treści na cztery kategorie[11]:

- rozrywkowe - przygotowywane w formie żartobliwej, wykorzystujące wizerunki postaci powszechnie znanych, postaci z popkultury lub postaci fikcyjne (np. popularne trendy w aplikacji TikTok ożywianie zdjęć osób zmarłych, postarzanie/odmładzanie wizerunku)[12];
- edukacyjne – wykorzystywanie wizerunku lub głosu znanej postaci w celu przedstawienia określonych treści edukacyjnych;
- dezinformacyjne – ukierunkowane na wzbudzeniu niepokoi społecznych lub wprowadzenia chaosu informacyjnego;
- dyskredytacyjne – generowane w celu zdyskredytowania osoby, grupy społecznej lub innego podmiotu[13].

Dwie pierwsze z powyższych kategorii, o ile nie zostały naruszone dobra osobiste ani prawa osób, których wizerunek został wykorzystany[14], czy też nie podają treści zafałszowanych i niezgodnych z prawem, same w sobie nie stanowią zagrożenia dla bezpieczeństwa informacyjnego, o tyle treści dezinformacyjne oraz dyskredytacyjne – zwłaszcza jeśli wymierzone są w dowódców, żołnierzy, funkcjonariuszy lub administracje publiczną – mogą prowadzić do zaburzenia prawidłowego przepływu informacji, wpływać na obniżenie morale, a nawet prowadzić do braku zaufania społecznego, paniki, czy nawet powstawania konfliktów.

Zagrożenia ze stosowania *deepfake*

Jak już wcześniej zostało wspomniane, produkty *deepfake* generowane w celu dezinformacyjnym lub dyskredytacyjnym mogą stanowić szczególnie istotne zagrożenie dla bezpieczeństwa informacyjnego państwa, zarówno w sferze cywilnej, jak i militarnej. Cechą charakterystyczną, determinującą fakt powstawania tego typu zagrożeń, jest nieprzewidywalność w zakresie rozpowszechniania oraz wykorzystywania wygenerowanych treści[15]. Z uwagi na fakt, że ich wykorzystywanie jest ściśle powiązane z publikowaniem oraz dalszym udostępnianiem w mediach społecznościowych, nie można przewidzieć na jaką skale, ani przez jaki czas dany obraz lub wideo będzie dostępne dla użytkowników, a możliwości ich wykorzystywania mogą przyjmować szeroki zakres, od kooperacji pozytywnej, po negatywną[16]. Jednocześnie, z uwagi

[11] I. Dąbrowska, Deepfake – nowy wymiar internetowej manipulacji, „Zarządzanie Mediami" 2020, t. 8(2), s. 96.
[12] https://www.tiktok.com/discover/o%C5%BCywianie-zdj%C4%99%C4%87-zmar%C5%82ych [dostęp: 09.11.2025].
[13] I. Dąbrowska, op.cit., s.96.
[14] Zobacz: K. M. Kiełpiński, op.cit., s. 91 oraz s. 94.
[15] Ibidem, s. 89.
[16] G. Pilarski, Cyberprzestrzeń - relacje w wojnie hybrydowej, Warszawa 2020, s. 28-30.

na ich wysoką (pozorną) wiarygodność oraz wykorzystywanie w mediach społecznościowych algorytmów tworzących tzw. „bańkę informacyjną”[17], mogą stanowić podstawę do wyrabiania sądów oraz opinii przez odbiorców, czy też stanowić umożliwiać prowadzenie innej działalności przestępczej (np. phishing, kradzież tożsamości)[18], a także stanowić wsparcie lub preludiuim do prowadzenia działań wojskowych[19].

Z powyższym łączy się również inne zagrożenie, jakim jest potęgowanie szumu informacyjnego. Z uwagi na fakt, że we współczesnym świecie ilość informacji jaka jest stale wytwarzania, przetwarzana oraz rozpowszechniana, znacząco wzrasta ryzyko nieprawidłowej interpretacji oraz weryfikacji pozyskiwanych informacji, co w konsekwencji prowadzić może do dezinformowania społeczeństwa, niepokoi społecznych, a w przypadku sytuacji kryzysowych, również do powstawania zbiorowej paniki, potęgując tym samym istniejące ryzyko[20]. Warto w tym miejscu podkreślić, że treści generowane w formie *deepfake*, nawet jeśli mają wyłącznie charakter rozrywkowy, w przypadku nieprawidłowej interpretacji przez odbiorcę lub celowe wykorzystanie z nieetycznych pobudek, dzięki swojej wizualnej formie, mogą być znacznie bardziej przystępne oraz uznane za wiarygodne, niż oficjalne oraz rzetelne komunikaty przekazywane przez podmioty państwowe[21].

Innym istotnym zagrożeniem wynikającym z *deepfake*, jest zacierania się świata rzeczywistego oraz wirtualnego[22]. Już dzisiaj można dostrzec przykłady, gdzie na podstawie importowanych do oprogramowania danych wejściowych, algorytmy tworzą niezwykle wiarygodne postacie oparte o charakterystyczne cechy człowieka, w taki sposób, że bez wnikliwego zbadania, nie można zweryfikować czy postać jest rzeczywista, czy też jedynie wytworem programu komputerowego[23]. Zagrożenie to, w kontekście rozpatrywanego bezpieczeństwa informacyjnego, może wywoływać długofalowe skutki, bowiem może znacząco wpływać na poprawność badania danych historycznych, czy statystycznych, których weryfikacja będzie w znaczącym stopniu utrudniona. Jednocześnie może to równie negatywnie wpływać na poczucie tożsamości i bezpieczeństwa

[17] P. Bączek, Dezinformacja jako zagrożenie dla bezpieczeństwa informacyjnego współczesnych społeczeństw, „Security Review” 2020, nr 4(17).
[18] O.Wasiuta, S. Wasiuta, op.cit., s. 22.
[19] G. Pilarski, Cyberprzestrzeń..., s. 30.
[20] D. Domalewska, Wielowymiarowość komunikacji w kontekście bezpieczeństwa. Komunikacja w sytuacjach kryzysowych i komunikacja strategiczna, Warszawa 2020, s. 134-135.
[21] O. Wasiuta, S. Wasiuta, op.cit., s. 26.
[22] Ł. Sarowski, Wpływ sztucznej inteligencji na infosferę – wybrane zagadnienia, „Tempus” 2024, nr 2, s. 151-152.
[23] Przykładami może być wykreowanie przez AI postaci Agaty Bąk, J. Kowalski, Postać stworzona przez sztuczną inteligencję wygrała plebiscyt Polska Press „Osobowość Roku”. „Weryfikujemy regulaminy” https://www.wirtualnemedia.pl/sztuczna-inteligencja-ai-zwyciestwo-osobowosc-roku-polsk a-press-manipulacja,7169658705974913a [dostęp: 09.11.2025], czy też eksperyment krakowskiego radia, w którym audycje prowadziła Sztuczna Inteligencja, wykorzystując technologię *deepfake,* w celu przeprowadzenia fikcyjnego wywiadu z Wisławą Szymborską, M. B. Lewkowicz, Wywiad z Szymborską wygenerowany przez AI. „Myślę, że by ją to rozbawiło”, https://cyberdefence24.pl/technologie/wywiad-z-szymborska-wygenerowany-przez-ai-mysle-ze-by-ja-to-rozbawilo [dostęp: 09.11.2025].

użytkowników bezpieczeństwa, zwłaszcza wśród dzieci i młodzieży który w sposób szczególny podatni są na manipulowanie percepcją, z uwagi na ograniczone umiejętności rozpoznawania oraz weryfikacji danych[24].

Omawiając zagrożenia dla bezpieczeństwa informacyjnego wynikające z wykorzystywania *deepfake*, nie można pominąć również niezwykle istotnego zagrożenia, jakim jest wykorzystywanie wygenerowanych treści na potrzeby działań realizowanych w ramach walki informacyjnej, dezinformacji, czy też operacji psychologicznych prowadzonych przez przeciwnika, których obiektem oddziaływania jest społeczeństwo własne lub sojusznicy. W tym zakresie *deepfake*, może być ukierunkowane zarówno na potrzeby zdyskredytowania pojedynczych decydentów lub dowódców (zarówno wobec podległych żołnierzy, przełożonych, czy też opinii publicznej), jak również pozycji państwa w strukturach międzynarodowych. Jednocześnie treści w ten sposób wygenerowane (np. obrazujące drastyczne sceny z pola walki), mogą wpływać na obniżenie morale żołnierzy oraz społeczeństwa, obniżając tym samym ich wolę walki[25]. Przykładem *deepfake*, który został przygotowany oraz rozpowszechniony w tym celu, jest spreparowane przez Federacje Rosyjską przemówienie prezydenta Ukrainy, w którym miał przekonywać ukraińskich żołnierzy do dezercji[26].

Wyzwania w zakresie weryfikacji i przeciwdziałania

Omówione powyższe zagrożenia generują dla państwa wyzwania w zakresie ich minimalizacji oraz zapobiegania niekorzystnym skutkom ich występowania.

Najistotniejszym z wyzwań, jest stworzenie możliwości weryfikacji treści pod kątem wykrycia wygenerowania materiału przy pomocy technologii *deepfake.* Z uwagi na fakt, że materiały te rozpowszechniane są głównie w cyberprzestrzeni, która stanowi stosunkowo trudne środowisko w zakresie wykrywania i zwalczania działań propagandowych[27], ich niezwykle precyzyjne wykonanie, bez wnikliwej analizy szczegółów, powoduje, że odróżnienie treści fałszywych od rzeczywistych, jest niemal niemożliwe. Zwłaszcza, że szczegóły te obejmują niewidoczne na pierwszy rzut oka niedoskonałości, jak niesynchroniczne mruganie, zmiany koloru skóry, zaburzenie perspektywy, nierealistyczne refleksy światła, nienaturalne deformacje na krawędzi twarzy,

[24] Więcej na ten temat mażna odnaleźć w: P. Galemba, M. Owsianka, Wpływ Internetu i nowych mediów na poczucie patriotyzmu i tożsamości narodowej, [w:] I. Machaj, J. Lubimow, Ł.Budzyński (red.), Młodzi dla bezpieczeństwa. Młodzi o bezpieczeństwie. Perspektywa Interdyscyplinarna, Gorzów Wielkopolski 2023, s. 121-135.

[25] Porównaj: G. Pilarski, Cyberprzestrzeń..., s. 30 oraz P. Galemba, Operacje Psychologiczne jako zagrożenie dla bezpieczeństwa państwa w XXI wieku, „Wiedza Obronna" 2025, vol 290, no 1, a także: O. Wasiuta, S. Wasiuta, op.cit., s. 22 oraz s. 26.

[26] P. Różyński, Deepfake idzie na wojnę, „Rzeczpospolita", 22 marca 2022, s. L2.

[27] Zobacz: G. Pilarski, Cyberspace as a Tool of Contemporary Propaganda, „Safety&Defense" 2020, t. 6.

ruch ust, etc.[28]. Dodatkową trudnością – jak zauważa K. Rzadkowska – na niewielkich ekranach szczegóły te są w zasadzie niewidoczne[29], co biorąc pod uwagę, że w głównej mierze treści *deepfake* rozpowszechniane są w mediach społecznościowych, przeglądanych zazwyczaj na telefonach, niemal całkowicie wyklucza samodzielną weryfikację ich rzetelności przez zwykłego użytkownika tych mediów.

Na potrzeby weryfikacji tworzone są specjalistyczne oprogramowania, oparte na sztucznej inteligencji, które – analogicznie jak w przypadku tworzenia tych treści – weryfikują je na podstawie zaimportowanych próbek obrazów, dźwięku i wideo. Jednakże, również one nie są w pełni wystarczające i wolne od ograniczeń i błędów. Głównymi czynnikami, które wpływają na ich ograniczenia są:

- ograniczona generalizacja i podatność na obejścia;
- wyzwania techniczne i wydajnościowe;
- kontekst i ograniczenia pozatechnologiczne;
- czynnik ludzki i poznawczy[30].

Kolejnym z wyzwań, jakie stoi przed państwem, jest konieczność zwiększenia poziomu edukacji medialnej oraz świadomości społeczeństwa w zakresie nie tylko zagrożeń wynikających z rozpowszechniania materiałów *deepfake* czy też braku weryfikacji treści w wiarygodnych źródłach, ale również w zakresie rozsądnego upubliczniania swoich prywatnych zdjęć i nagrań w przestrzeni informacyjnej (np. na portalach społecznościowych), które później mogą posłużyć jako materiał na potrzeby przygotowania produktów *deepfake*, bazujących na tych materiałach, w celu uzyskania przez twórcę określonych nieetycznych celów (finansowych lub w dyskredytacji). Jednocześnie, drugim z zagadnień, które powinno być elementem edukacji społeczeństwa, powinno być uświadamianie w zakresie rozsądnego wykorzystywania narzędzi opartych na sztucznej inteligencji, nie tylko generujących obrazy, czy filmy, ale również znacznie prostsze, generujące odpowiedzi tekstowe (np., Chat GPT). Wynika to przede wszystkim z faktu, że jest to technologia „samoucząca się", na bazie wprowadzanych danych, a każde jej wykorzystanie, prowadzi do jej dalszego samorozwoju. W konsekwencji prowadzi to do stworzenia błędnego koła, w którym algorytmy weryfikujące są zacofane względem możliwości generowania treści.

Ostatnim z wyzwań w zakresie minimalizacji negatywnych skutków wykorzystywania *deepfake*, jest konieczność wprowadzenia stosownych regulacji prawnych. Jak już zostało wspomniane wcześniej, zjawisko to jest stosunkowo nowe oraz stale się zmieniające i rozwijające, a co za tym idzie

[28] Zobacz: K. Basaj, Czym jest deepfake?, „Deepfake – wydanie specjalne Biuletynu ACKS", 2021, nr 2, s. 9 oraz M. Chodyk, Z. Ciekanowski, Analiza treści audiowizualnych jako element systemów bezpieczeństwa państwa, Biała Podlaska 2024, s. 114-115.
[29] K. Rzadkowska, Deepfake w komunikacji, „Deepfake – wydanie specjalne Biuletynu ACKS", 2021, nr 2, s. 12.
[30] M. Chodyk, Z. Ciekanowski, op.cit., s. 116-119.

obecnie obowiązujące regulacje prawne nie są wystarczające w tym zakresie[31]. Zmiany te powinny objąć przede wszystkim bezwzględną konieczność oznaczania wszelkiego rodzaju treści wygenerowanych przez sztuczną inteligencję, nie tylko w zakresie *deepfake*, czy grafik/nagrań umieszczanych w przestrzeni internetowej, ale również rozpowszechnianych w formie plakatów czy spotów reklamowych, a pod rozwagę powinno zostać wzięte także całkowite zakazanie wykorzystywanie tego rodzaju rozwiązań w mediach masowego przekazu[32].

Zakończenie

Podsumowując powyższe rozważania nad istotą *deepfake* oraz wynikających z nich zagrożeń i wyzwań dla bezpieczeństwa informacyjnego państwa należy zauważyć, że jest to stosunkowo nowe zjawisko, a pełnia jego możliwości nie jest jeszcze znana. Znaczący wzrost intensywności wykorzystywania tej technologii wynika w głównej mierze z niewielkich kosztów przygotowywania materiałów, jak również z wykorzystywanie do tego celu prostych oraz intuicyjnych programów komputerowych, do których obsługi nie są wymagane specjalistyczne urządzenia. Wraz z rozwojem możliwości generowania materiałów *deepfake*, wzrasta również ryzyko występowania związanych z nimi zagrożeń, m.in. w zakresie powstawania chaosu informacyjnego, potęgowanie napięć społecznych, zacierania granicy pomiędzy światem realnym i wirtualnym, a także wzrost efektywności działań realizowanych w ramach walki informacyjnej oraz oddziaływani psychologicznego, ukierunkowane na dyskredytacje przywódców i dowódców, grup społecznych oraz pojedynczych osób, obniżenie rangi państwa na arenie międzynarodowej, czy też wpływanie na morale żołnierzy, funkcjonariuszy i społeczeństwa, potęgując jednocześnie zjawisko dezinformacji.

Wyzwaniami w tym zakresie są natomiast konieczność zwiększenia możliwości weryfikacji treści typu *deepfake*, poprzez wykorzystywanie specjalistycznego oprogramowania opartego na algorytmach umożliwiających analizę treści, których odróżnienie od rzeczywistych obrazów jest niemal niemożliwe dla standardowego użytkownika Internetu. Kolejnym wyzwaniem jest wprowadzenia regulacji prawnych ukierunkowanych na ograniczenie nieuprawnionego wykorzystywania wizerunku osób, a także wykorzystywania nie tylko technologii *deepfake*, ale także szeroko pojmowanej sztucznej inteligencji, zwłaszcza w przestrzeni medialnej.

Ostatnim wyzwaniem jest konieczność zwiększenie poziomu świadomości oraz edukacja społeczeństwa w zakresie racjonalnego upubliczniania swojego wizerunku na portalach społecznościowych, a także ograniczanie bezmyślnego

[31] K. M. Kiełpiński, op.cit., s. 89.
[32] Przykładem może być Australia, w której wprowadzono ograniczenia w zakresie wykorzystywania *deepfake* oraz sankcje. Zob.: O. Wasiuta, S. Wasiuta, op.cit., s. 23.

wykorzystywania technologii *deepfake,* której z uwagi na mechanizmy samouczące, permanentnie zwiększają swoje możliwości w zakresie kreowania coraz wiarygodniejszych i trudniejszych do wykrycia fałszerstw.

Reasumując, powyższe rozważania pozwoliły udzielić odpowiedzi na sformułowane we wprowadzeniu pytanie, a tym samym rozwiązać postawiony problem badawczy oraz wskazać główne zagrożenia i wyzwania dla bezpieczeństwa informacyjnego państwa, związane z wykorzystywaniem technologii *deepfake*.

Na koniec warto jeszcze podkreślić, że zagrożenia i wyzwania wynikające ze stosowania technologii *deepfake* ulegają permanentnym zmianom, a ich katalog oraz intensywność będą stale wzrastały z uwagi na fakt, że sztuczna inteligencja, na której jest oparta, wchodzi w kolejne fazy rozwoju, stwarzając nowe możliwości generowania, rozpowszechniania i wykrywania fałszywych treści wykorzystujących wizerunek oraz głos człowieka. Państwa (oraz organizacje międzynarodowe) będą stawać natomiast nie tylko przed nowymi wyzwaniami z nich wynikającymi, ale także przed dylematem: zezwolić nieograniczony postęp technologiczny, czy też wprowadzić ograniczenia umożliwiające zachowanie bezpieczeństwa?

Bibliografia

Basaj K., Czym jest deepfake?, „Deepfake – wydanie specjalne Biuletynu ACKS”, 2021, nr 2.

Bączek P. Dezinformacja jako zagrożenie dla bezpieczeństwa informacyjnego współczesnych społeczeństw, „Security Review” 2020, nr 4(17).

Chodyka M., Ciekanowski Z., Analiza treści audiowizualnych jako element systemów bezpieczeństwa państwa, Biała Podlaska 2024.

Czerski M. W., Filter bubbles jako narzędzie dezinformacji i manipulacji, „Humanities and Culture Studies” 2022, vol. 3, no.2.

Dąbrowska I., Deepfake – nowy wymiar internetowej manipulacji, „Zarządzanie Mediami” 2020, t. 8(2).

Domalewska D., Wielowymiarowość komunikacji w kontekście bezpieczeństwa. Komunikacja w sytuacjach kryzysowych i komunikacja strategiczna, Warszawa 2020.

Galemba P., Operacje Psychologiczne jako zagrożenie dla bezpieczeństwa państwa w XXI wieku, „Wiedza Obronna" 2025, vol 290, no 1.

Galemba P., Owsianka M., Wpływ Internetu i nowych mediów na poczucie patriotyzmu i tożsamości narodowej [w:] I. Machaj, J. Lubimow, Ł.Budzyński (red.), Młodzi dla bezpieczeństwa. Młodzi o bezpieczeństwie. Perspektywa Interdyscyplinarna, Gorzów Wielkopolski 2023.

https://www.tiktok.com/discover/o%C5%BCywianie-zdj%C4%99%C4%87-zmar%C5%82ych.

Janczak J., Nowak A., Bezpieczeństwo informacyjne. Wybrane problemy. Warszawa 2013.

Kacała T., Dezinformacja i propaganda w kontekście zagrożeń dla bezpieczeństwa państwa, „Przegląd Prawa Konstytucyjnego” 2015, nr 2(24).

Kai-Fu Lee, Inteligencja sztuczna rewolucja prawdziwa. Chiny, USA i przyszłość świata, Poznań 2019.

Kai-Fu Lee, Chen Quifan, AI 2041: Ten visions for our future, London 2024.

Kiełpiński K. M., Deepfake jako narzędzie do przekazywania informacji fałszywej i domniemanej. Analiza prawnokarna i cybernetyczna, „Kwartalnik Krajowej Szkoły Sądownictwa i Prokuratury", 2023, z. 3(51).

Kowalski J., Postać stworzona przez sztuczną inteligencję wygrała plebiscyt Polska Press „Osobowość Roku". „Weryfikujemy regulaminy" https://www.Wirtualnemedia.pl/sztuczna-inteligencja-ai-zwyciestwo-osobowosc-roku-polska-press-manipulacja,7169658705974913a.

Kwećka R., Bezpieczeństwo informacyjne, [w:] J. Pawłowski (red.), Podstawy bezpieczeństwa narodowego (państwa), Warszawa 2017.

K. Chałubińska-Jentkiewicz (red.), Leksykon cyberbepieczeństwa, Warszawa 2024.

Lewkowicz M.B., Wywiad z Szymborską wygenerowany przez AI. „Myślę, że by ją to rozbawiło",https://cyberdefence24.pl/technologie/wywiad-z-szymborska-wygenerowany-przez-ai-mysle-ze-by-ja-to-rozbawilo.

Modrzejewski Z., Działania psychologiczno-propagandowe w konfliktach zbrojnych, „Kwartalnik Bellona" 2017, nr 1.

Mularz K., Deepfake jako generator fikcyjnych treści i rzeczywistej nieufności, „Dydaktyka Informatyki" 2024, nr 19.

Nowak E., Nowak M., Zarys teorii bezpieczeństwa narodowego, Warszawa 2011.

Piecuch A. , Życie SMART - ma swoją cenę, „Dydaktyka Informatyki" 2020, nr 15.

Pilarski G., Cyberprzestrzeń – relacje w wojnie hybrydowej, Warszawa 2020.

Pilarski G., Cyberspace as a Tool of Contemporary Propaganda, „Safety&Defense", 2020, t. 6.

Różyński P., Deepfake idzie na wojnę, „Rzeczpospolita", 22 marca 2022.

Rzadkowska K., Deepfake w komunikacji, „Deepfake – wydanie specjalne Biuletynu ACKS" 2021, nr 2.

Sarowski Ł., Wpływ sztucznej inteligencji na infosferę – wybrane zagadnienia, „Tempus" 2024, nr 2.

Słownik terminów z zakresu bezpieczeństwa, J. Pwałowski, B. Zdrodowski, M. Kuliczkowski, (red.), Toruń 2020.

Wasiuta O., Wasiuta S., Deepfake jako skomplikowana i głęboko fałszywa rzeczywistość, „Annales Univeristatis Paedagogicae Cracoviensis. Studia de Securitate" 2019, nr 9(3)